

# Research Topics on Web Intelligence and Communities

Rajendra Akerkar<sup>a</sup>, Pierre Maret<sup>b,\*</sup> and Laurent Vercoeur<sup>c</sup>

<sup>a</sup> *Western Norway Research Institute (Vestlandsforskning) Sogndal, NO-6851, Norway*

*E-mail: rak@vestforsk.no*

<sup>b</sup> *Université de Lyon, CNRS, UMR5516, Laboratoire Hubert Curien, F-42000, Saint-Etienne, France,*

*E-mail: pierre.maret@univ-st-etienne.fr*

<sup>c</sup> *Institut National de Sciences Appliquées de Rouen. 76801 Saint-Etienne-du-Rouvray Cedex, France*

*E-mail: laurent.vercoeur@insa-rouen.fr*

**Abstract.** Web Intelligence deals with intelligent methods and information & communication technologies that are integrated to enhance different web-based applications. Communities are popular, particularly on the World Wide Web, as a means for like-minded individuals to pursue common goals. Communities appear as a first-class object in the areas of web intelligence and agent technologies, as well as a crucial crossroads of several sub-domains (i.e. user modeling, protocols, data management, data mining, content modeling, etc.). These sub-domains impact the nature of the communities and the applications which are related to them. The use of Web Intelligence and communities is discussed together with ways in which a wide range of research is benefiting this area for the long-term.

**Keywords:** Distributed Artificial Intelligence, Web intelligence, web communities, multi-agent, retrieval, social networks

## 1. Introduction

The rapid growth of Web technology has made the World Wide Web an important and popular application platform for disseminating and searching information as well as conducting business. This growth gave a way to the development of ever smarter approaches to extract patterns and build knowledge with the aid of artificial intelligence techniques. These techniques have been used, together with information technology, in a wide range of applications. This is where semantics, social network analysis, web structure, content, usage, and other aspects have already been and will increasingly keep being included in many application domains. The web provides rich medium for communication, which goes far beyond the conventional communication media.

To keep up-to-date in the research areas of Web Intelligence & Communities is fundamental to further

contribute towards the understanding of how the Web can improve to our everyday life.

## 2. Overview of the research field

### 2.1. Definitions

Web Intelligence [1] consists of a multidisciplinary area dealing with exploiting data and services over the Web, to create new data and services using both Information and Communication Technologies (ICT) and Artificial Intelligence (AI) techniques. Different objectives have been followed, and different approaches and technologies have been used by researchers and practitioners over the years. We can mention concepts such as Web information repositories, Web user behavior analysis, Web content and structure mining, social network analysis, the semantic Web. In addition more general concepts such as Knowledge Discovery from Databases, Multi-Agent Systems, Machine Learning,

---

\*Corresponding author. E-mail: pierre.maret@univ-st-etienne.fr

Knowledge Representation, and Distributed Systems are some keys to understand the fundamentals of intelligent web.

A Web Community can be defined as a group of individuals sharing common interests and making use of electronic forms of communication for exchanges on the Web. Thus, a web community is not necessarily related to a task, rather to a topic and to some knowledge. Since it can also be related to a task, one may say that it is more general than a virtual enterprise and a virtual team. In particular, the Web has several familiar explicitly and implicitly defined communities. Explicit communities are the communities that are available to be identified easily on the Web. Kumar et al. [15] discussed the example of an explicit community of Web users interested in Porsche Boxster cars, such as the Porsche newsgroup, or resource collections in directories in search engines, such as the Yahoo directory. Explicit communities are easy to be identified and analysts can merely use manual method to find an enterprise's explicit communities by browsing the enterprise's newsgroup, or the category in which the enterprise falls into in the directory like Yahoo on the Internet.

Implicit communities are comparatively more complex to find using manual browsing method. According to Kumar, implicit communities refer to the distributed, ad-hoc and random content-creation related to the common interests on the Internet [15]. The pages often have links to each other, but the common interests of implicit communities are sometimes too narrow and detailed for the resource pages or the directories to develop explicit listings for them. As a result, it is more difficult to find the implicit communities of an enterprise. In identifying the explicit and implicit communities, it is often assumed that the content pages created by these communities would provide hypertext links back to the enterprise's homepage for reference [16].

## 2.2. Models and repositories

With respect to knowledge representation and repositories, areas such as logic, ontology, and computation are vital in order to support the basic structure evolving from a Web of data to a Web of knowledge [2]. Besides, once knowledge is mined from the web data, different standards have been developed to store and manage the different patterns extracted from the content. These repositories have been developed for use in Multidimensional Analysis architectures. This is where Extraction, Transformation, and Loading from web-

based resources, Data Web-based Meta-data Modeling, OLAP queries, and its visualization have been broadly studied [3]. Various web mining applications, such as Web User Behavior, Content of Web Sites, and the Analysis of the Web as a graph have been considered in the field of Web Intelligence, Web Mining, Machine Learning, Information Retrieval, and Artificial Intelligence communities in international conferences and journals.

The web usage mining researchers have extensively investigated web usability and usefulness considerations, for instance, helping the web user to obtain information [2]. Moreover, the content of a given web site has formed the focus for Web Content Mining researchers [1]. The structure, representation, and its analysis have been considered as part of Web structure mining [4] and the information retrieval [5]. In order to offer Web data in suitable formats, Web logs, the Web-site contents, and the Hyperlink Structure of the Web, have been considered as the main source of information. Privacy issues, such as using invasive tools to identify the users, and social network analysis where the user's contacts are exposed, have been the focus of further developments in privacy preserving intelligent web applications [6,7,8].

Early research on web structure has led to various ranking algorithms that are now used in the analysis on how communities are formed. This includes the HITS algorithm, where authorities and hubs are established [11]. The web content is being developed by its users in web blogging, virtual communities, on-line knowledge communities, web forums, microblogging, on-line encyclopedias, and social network applications. This enables the storage and generation of linked and structured information, that can be associated with text messages and multimedia information such as images and videos. Mining into these various multimedia contents provides insights on users and contributes to determine user profiles.

Recently acclaimed research areas in the field of web intelligence and communities are the social networks and in web communities' analysis [9,10]. Furthermore, it is also focused on the growth of the semantic Web. The key intention is to give a Web of eloquent meaning. There are various aspects of knowledge representation such as computational linguistics, which have contributed to its development [12]. Several standards for meta-data processing such as the Resource Description Framework (RDF), Web Ontology Language (OWL), semantically interlinked on-line communities (SIOC), and social network representations

of RDF, such as Friend of a Friend (FOAF), have been accepted as the standards to semantics concerns in the Web 3.0 [17].

Socially enabled Web information [13] search is a new phenomenon facilitated by recent Web technologies. This collaborative social search involves finding specific people in your network who have the knowledge you're looking for or finding relevant information based on one's social network. People in social groups can provide solutions, pointers to databases or other people (meta-knowledge), validation and legitimization of ideas, can serve as memory aids and help with problem reformulation. In [14], a Conversational Search and Recommendation system that involves finding relevant information based on social interactions and feedback along with augmented agent based recommendations is introduced.

### 3. Papers of the special issue

The 5 papers of this special issue are extensions of papers selected for the *Web Intelligence & Communities* workshop of years 2011 and 2012, hosted respectively at the "Web Intelligence and Intelligent Agent Technologies" and at the "World Wide Web" conferences. Twenty papers were presented during these workshops. The extended papers went through distinct review process. of 5 of them are presented in this special issue. Many concerns should be addressed in the Web Intelligence and Community area; the papers selected for this special issue address some of these concerns.

#### Sebastian Rios and Roberto Silva

*A New Dissimilarity Measure for Online Social Networks Moderation.*

Authors present a Dissimilarity Measure for Automate Moderation in Online Social Networks. The authors proposes to combine web usage mining (WUM) techniques with latent context semantics to address of the issue of moderating online social networks. Their approach proposes to automate this moderation task and to detect detect users that need to be moderated by the community administrators.

#### Reda Yaich, Olivier Boissier, Gauthier Picard, and Philippe Jaillon

*Adaptiveness and Social-Compliance in Trust Management within Virtual Communities.*

In the paper, a trust for the management of trust in virtual communities is proposed, based on policies in a multi-agent environment. Thanks to meta-rules, trust policies of agents can adapt automatically in the virtual community.

#### Ioan Alfred Letia and Octavian Pop

*Towards Adaptive Normative Systems for Communities of Agents.*

Authors propose a formal model using description logic for adaptive norms enforcement in virtual communities. A multi-agent based virtual community scenario is used to illustrate the applicability of the approach.

#### Sebastian Rios, Felipe Aguilera, Francisco Bustos, Tope Omitola, and Nigel Shadbolt

*Leveraging Social Network Analysis with Topic Models and the Semantic Web (extended).*

This article presents an extension SIOC ontology to associate entities of a social network with the relationships extracted from the content shared on the platform. A text mining approach based on the Latent Dirichlet Allocation model is used to analyse contents. Authors propose then to represent community's purposes and policies in their model.

This collection of papers comprises an illustrative piece of the contribution of the Web Intelligence field for Web Communities.

Although several aspects are still under investigation, and others need to be taken into consideration, Web Intelligence is situated to play a vital role in the next generation of Web applications. Without a doubt, the web communities field will remain to create a strong impact in the future and major challenge of this research in the next decade will be Web Intelligence centric, focusing on how intelligently make the best use of the widely available web connectivity.

### 4. List of reviewers

The invited editors of this special issue are grateful to reviewers who spent time reading and evaluating the papers. Their recommendations have been appreciated by all authors.

Reviewers: Amblard Frederic, Badica Costin, van Beijnum Bert-Jan, Bothorel Cecile, Camacho David,

Champin Pierre-Antoine, Ding Ying, Elmorr Christo, Langeron Christine, Lopez Guillaume, Picard Gauthier, Piunti Michele, Ribiere Myriam, Stan Johann, Stratulat Tiberiu, Ting I-Hsien, Yaich Reda.

## References

- [1] Akerkar, R. and Lingras, P. *Building an Intelligent Web: Theory & Practice*. Jones and Bartlett, Sudbury, MA. 2008.
- [2] Velasquez, J.D., Palade, V. *Adaptive Web Sites: A Knowledge Extraction from Web Data Approach*. IOS Press, Amsterdam (2008)
- [3] Rebolledo, V.L., Velásquez, J.D. *A platform for extracting and storing web data*. In: Velásquez, J.D., Rios, S.A., Howlett, R.J., Jain, L.C. (eds.) *Knowledge-Based and Intelligent Information and Engineering Systems*. LNCS (LNAI), vol. 5711, pp. 843–850.
- [4] Liu, B. 2007. *Web Data Mining: Exploring Hyperlinks, Content and Usage Data*, 1st edn. Springer, Heidelberg.
- [5] Baeza-Yates, R.A., Ribeiro-Neto, B. 1999. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston.
- [6] Agrawal, R., Srikant, R. 2000. *Privacy-preserving data mining*. SIGMOD Rec. 29(2), 439–450.
- [7] Xu, Y., Wang, K., Zhang, B., Chen, Z. 2007. *Privacy-enhancing personalized web search*. In: WWW 2007: Proceedings of the 16th international conference on World Wide Web, pp. 591–600. ACM Press, New York
- [8] Krupa Y., Vercouter L., 2012. *Handling privacy as contextual integrity in decentralized virtual communities: The PrivaCIAS framework*. In : Web Intelligence and Agent Systems, vol 10 (1), pp. 105-116, IOS Press
- [9] Golbeck, J., Rothstein, M. 2008. *Linking social networks on the web with foaf: a semantic web case study*. In: AAAI 2008: Proceedings of the 23rd national conference on Artificial intelligence, pp. 1138–1143. AAAI Press, Menlo Park.
- [10] Akerkar, R. and Aaberge, T. 2011. *Semantically linking virtual communities*. (Eds. Christo El Morr & Pierre Maret) Virtual Community Building and the Information Society: Current and Future Directions, pp. 192-207, IGI Global Publishers.
- [11] Kleinberg, J.M. 1999. *Authoritative sources in a hyperlinked environment*. J. ACM 46(5), 604–632.
- [12] Shadbolt, N., Berners-Lee, T., Hall, W. 2006. *The semantic web revisited*. IEEE Intelligent Systems 21(3), 96–101.
- [13] Horowitz, D. and Kamvar, S.D. 2010. *The Anatomy of a Large-Scale Social Search Engine*. Proceedings of ACM WWW 2010.
- [14] Venkatesh, A., Sahay, S., Ram, A. 2010. *Cobot: Real Time Multi User Conversational Search and Recommendations, Recommender Systems and The Social Web* at ACM RecSys
- [15] Kumar, R., Raghavan, P., Rajagopalan, S., and Tomkins, A. 1998. *Trawling the Web for Emerging Cyber-communities*, Proceedings of the 8th International World Wide Web Conference.
- [16] Reid, E. O. F. 2003. *Identifying a Company's Non-Customer Online Communities: a Proto-typology*. Proceedings of the Hawaii International Conference on System Sciences, Big Island, Hawaii.
- [17] Story, H., Blin, R., Subercaze, J., Gravier, C., Maret, P. 2012. *Turning a Web 2.0 social network into a Web 3.0, distributed, and secured social web application*. WWW (Companion Volume) 2012. 417-420.